AI Profiler

Nsight Systems and TensorBoard

OpenACC
More Science, Less Programming

Learning Objectives

You will learn to optimize AI applications using the NVIDIA® Nsight™ Systems and PyTorch Profiler with TensorBoard. Throughout the labs, you will:

- Learn how to profile an application using **NVIDIA Nsight Systems.**
- Use **PyTorch Profiler** to profile an application and visualize on **TensorBoard.**
- Interpret the Timeline provided by **NVIDIA Nsight Systems.**
- Use **TensorBoard** to understand the application's use of the system resources.
- Identify performance problems in applications and apply optimization strategies.
- Confirm the performance improvement gained from the optimizations.

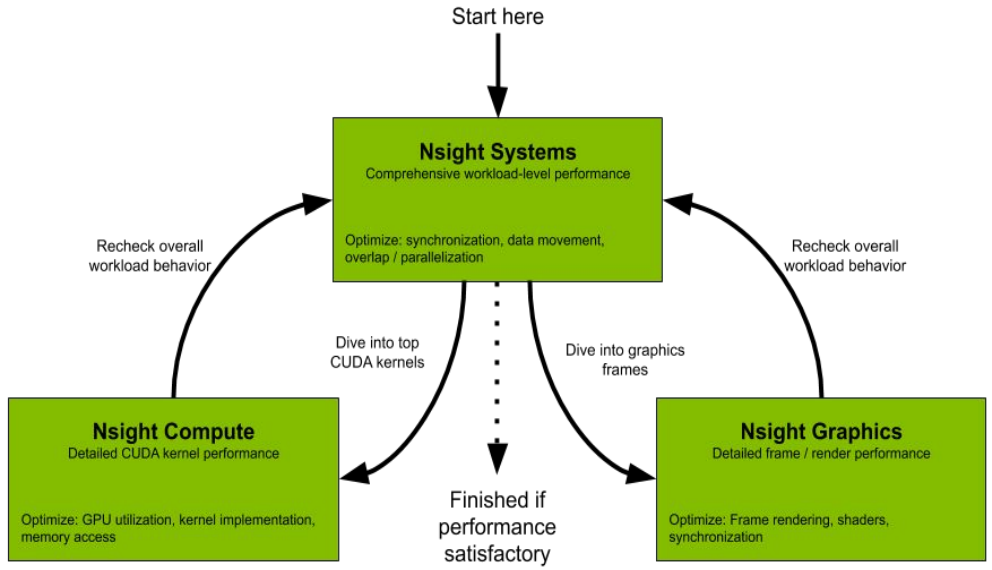
Introduction

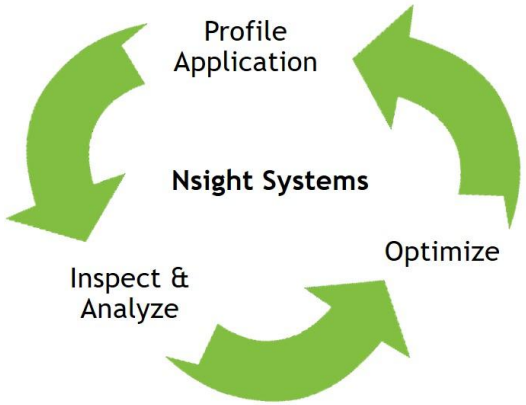
What is profiling?

- Profiling is the first step in optimizing and tuning your application.
- Profiling an application helps you understand where most of the execution time is spent.
- With profiling, you gain an understanding of the application's performance characteristics and identify parts of the code that present opportunities for improvement.
- Profiling enables you to find hotspots and bottlenecks in your application so you can decide where to focus your optimization efforts.

Nsight Systems

- **NVIDIA Nsight Systems** offers system-wide performance analysis in order to visualize application's algorithms.
- It helps identify optimization opportunities and improve the performance of applications running on a system consisting of multiple CPUs and GPUs.





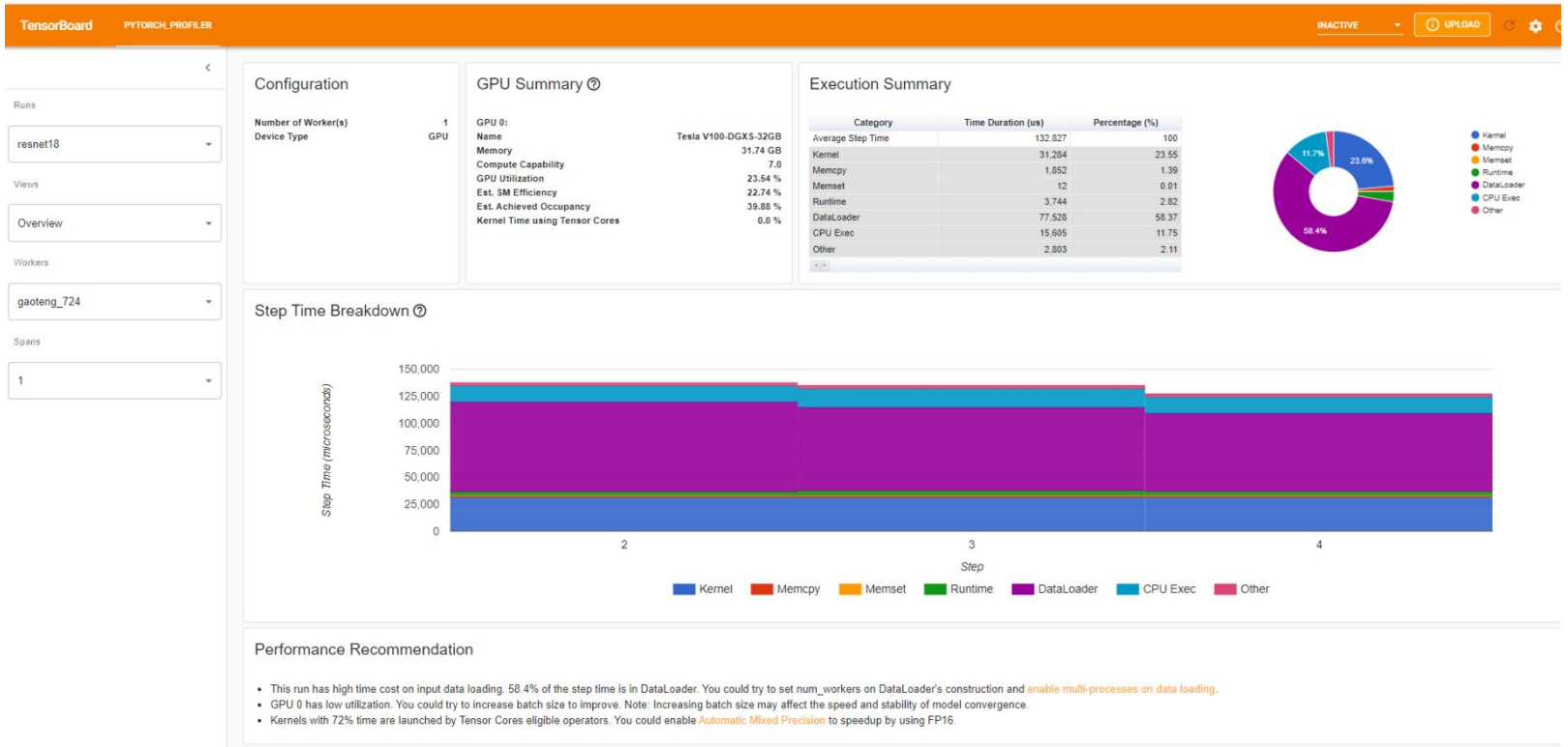
- **It is an iterative process; with 3 main steps**

✔ Profile the application

✔ Inspect and analyze the profile to identify any bottlenecks

✔ Optimize the application to address the bottlenecks

PyTorch Profiler with TensorBoard



- The **TensorBoard** is the visualization toolkit for TensorFlow. It provides tooling needed for tracking and visualizing performance metrics in a machine learning workflow.

- The **PyTorch Profiler** tool enables the profiling of deep neural networks (DNN) training program through the collection of performance metrics that include execution time, memory costs, stack traces, device Kernel, etc.

Hands-On

The Lab has 2 sections:

Section 1

- Profiling using NVIDIA Nsight Systems

Section 2

- PyTorch Profiler with TensorBoard plugin.

Both sections focus on steps to optimizing deep neural network training program using a **PyTorch mnist** program.

Minimum requirements to run the lab :

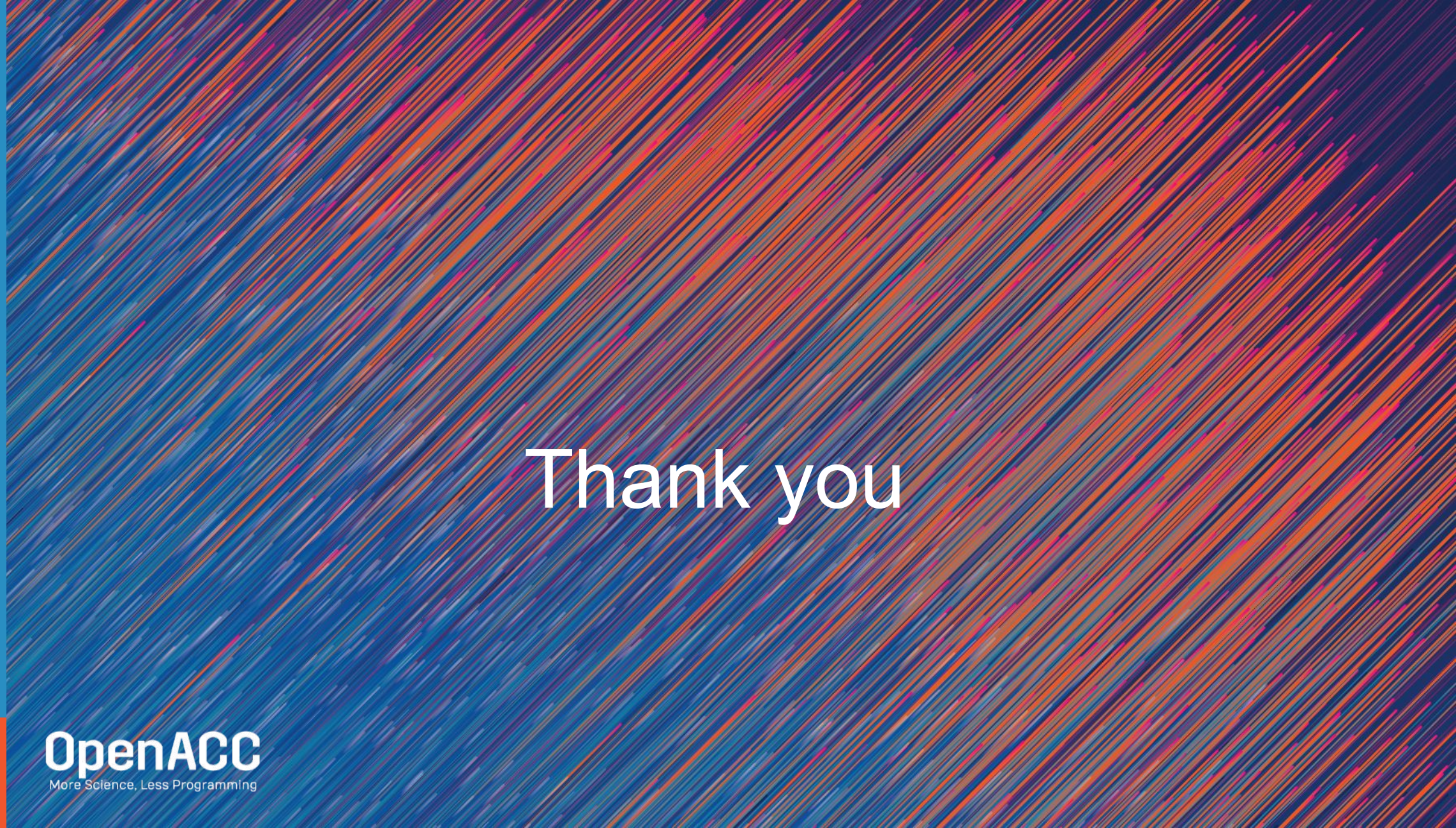
- NVIDIA Nsight Systems 2022.2.1 GUI and CLI
- Workstation or Local machine with GPU
- Docker or Singularity PyTorch container and TensorBoard plugin

Resources and Links

- Additional resources
  - NVIDIA Nsight Systems
  - NVIDIA Nsight Compute
  - Open Hackathons technical resource page
  - Open Hackathons GitHub Repository

- Join the OpenACC and Hackathons Slack channel

- Licensing


Thank you
OpenACC
More Science, Less Programming